Research article

Prognostic value of partial-EMT-related genes in head and neck squamous cell carcinoma by a bioinformatic analysis

Satoru Kisoda[1], Wenhua Shao[1], Natsumi Fujiwara[2], Yasuhiro Mouri[3], Takaaki Tsunematsu[1], Shengjian Jin[1], Rieko Arakaki[1], Naozumi Ishimaru[1], Yasusei Kudo[1,*]

[1]Department of Oral Molecular Pathology, [2]Department of Oral Health Care Promotion, and [3]Department of Biochemistry, Tokushima University Graduate School of Biomedical Sciences, Tokushima 770-8504, Japan.

Short running title: Prognostic value of p-EMT-related genes in HNSCC

[*]**Correspondence:** Yasusei Kudo, Department of Oral Molecular Pathology, Tokushima University Graduate School of Biomedical Sciences, 3-18-15 Kuramoto, Tokushima 770-8504, Japan. E-mail address: yasusei@tokushima-u.ac.jp

Funding information: This work was supported in part by grants-in-aid from the Ministry of Education, Science and Culture of Japan [grant numbers 19H03853].

Abstract

Objective: Recent studies have revealed that the ability of cancer cells to undergo intermediate state of EMT, partial EMT (p-EMT) poses a higher metastatic risk rather than complete EMT. Here we examined the prognostic value of p-EMT-related genes in head and neck squamous cell carcinoma (HNSCC) by bioinformatic approaches.

Materials and Methods: We used RNA-seq data of 519 primary HNSCC cases obtained from TCGA database. We compared the expression of p-EMT-related genes in HNSCC tissues with normal tissues. We evaluated the prognostic value of p-EMT-related genes in HNSCC cases by Log-rank test. We examined the expression of p-EMT-, EMT-, and epithelial differentiation-related genes by qPCR.

Results: Among p-EMT-related genes that were highly expressed in HNSCC cases, high expression of SERPINE1, ITGA5, TGFBI, P4HA2, CDH13, and LAMC2 was significantly correlated with poor survival of HNSCC patients. By gene expression pattern, HNSCC cell lines were classified into three groups; epithelial phenotype, EMT-phenotype, and p-EMT phenotype.

Conclusions: Our findings suggest that p-EMT program may be involved in poor prognosis of HNSCC. *SERPINE1, ITGA5, TGFBI, P4HA2, CDH13*, and *LAMC2* can be used for a prognostic marker. Moreover, HNSCC cells with p-EMT phenotype can be a useful model for investigating a nature of p-EMT.

KEYWORDS

partial-epithelial-to-mesenchymal transition; head and neck squamous cell carcinoma; prognosis

1. INTRODUCTION

Like most epithelial cancers, HNSCC develops through the accumulation of multiple genetic and epigenetic alterations in a multistep process. In cancer progression, it is known that epithelial-to-mesenchymal transition (EMT) in cancer cells is associated with invasion, metastasis, stemness, and resistance of therapy (Nieto et al, 2016). In the process of EMT, cancer cells lose their epithelial features, such as intercellular connections and E-cadherin expression and gain mesenchymal features. EMT is not binary process. It recently has been shown the intermediate state of EMT, partial EMT (p-EMT). Cancer cells with p-EMT state act like cancer cells with mesenchymal state, but they don't lose their epithelial feature completely (Pastushenko et al, 2018). Recent evidences showed that p-EMT program is involved in tumor progression and metastasis (Bednarz-Knoll et al, 2012; Saitoh, 2018). Moreover, *in vivo* study has revealed that p-EMT promotes tumor cell migration and formation of circulating tumor cell clusters (Aiello et al, 2018). Cells with p-EMT program have been referred to as ''metastable'', reflecting the flexibility of these cells to induce or reverse the EMT process (Lee et al, 2014; Tam et al, 2013). Indeed, it has been demonstrated that reversion of EMT is essential for disseminated tumor cells to proliferate and form metastases *in vivo* (Ocaña et al, 2012; Tsai et al, 2012). In similar to other types of cancer, p-EMT program is dynamic, invasive, and potentially responsive to tumor microenvironment cues in HNSCC, demonstrated by *in vivo* profiles and *in vitro* functional data (Puram et al, 2017). Single-cell RNA-seq of HNSCC cells identified p-EMT-related extracellular matrix genes (Puram et al, 2017). However, the nature of p-EMT and its prognostic value are still unknown. In this study, therefore, we examined the prognostic value of p-EMT-related genes by using public RNA-Seq data of primary HNSCC cases.

2. MATERIALS AND METHODS

Data analysis. RNA-sequencing data from 43 normal samples and 519 HNSCC samples released by TCGA (https://tcga-data.nci.nih.gov/tcga/). There was >10 year follow up time available for all data points. The RNASeqV2 data from TCGA was reanalyzed by R (ver. 3.6.1) and Subio Platform (ver. 1.22, Subio inc., Japan). For

investigating prognostic value of p-EMT-related genes in HNSCC cases, we divided the cases into two groups, "low" and "high", based on the median expression level of each p-EMT-related gene. Then, we compared the survival rate of "low" group with "high" group by log-rank test. Survival curve was made by Survminer package, and the heatmap was made by ggplot2 or pheatmap package. For log-rank test, a *P*-value of <0.02 was considered significant. For other analyses, a *P*-value of <0.05 was considered significant.

Cell culture. HSC2, HSC3, and HSC4 cells were provided by Japanese Collection of Research Bioresources Cell Bank. HOC313, HOC621, HOC719PE and HOC719NE cells were provided from Prof. Kamata (Hiroshima University). We previously found that HOC719NE and HOC313 have EMT features including loss of E-cadherin expression (Nguyen et al, 2013; Yokoyama et al, 2001). MSCC-inv1 cells were previously established in our laboratory. MSCC-inv1 cells were isolated as a high invasive clone by using *in vitro* invasion assay from MSCC-1 cells that were established from lymph node metastasis of gingival cancer (Kudo et al, 2003; Kudo et al, 2004)

qPCR. Total RNA was isolated from cultures of confluent cells using the RNeasy Mini Kit (Qiagen). Preparations were quantified and their purity was determined by standard spectrophotometric methods. cDNA was synthesized from 1 μg total RNA according to the PrimeScript II reverse transcriptase (Takara Bio Inc.). Expression levels of p-EMT, EMT and epithelial differentiation genes were determined using a LightCycler 96 system (Roche) with TB Green Premix Ex Taq II (Takara Bio Inc.). List of primers is shown in Supplemental Table 1. Relative mRNA expression of each transcript was normalized against GAPDH mRNA.

3. RESULTS

Recent single cell transcriptome analysis has identified several genes that are involved in p-EMT program (Puram et al, 2017). We focused on the representative p-EMT-related genes; 15 common p-EMT-related genes (*SERPINE1, TGFBI, MMP10,*

LAMC2, *P4HA2*, *PDPN*, *ITGA5*, *LAMA3*, *CDH13*, *TNC*, *MMP2*, *EMP3*, *INHBA*, *LAMB3*, and *VIM*) and 10 valuable p-EMT-related genes (*THBS2*, *CXCL13*, *FN1*, *MMP3*, *MMP9*, *RAB25*, *MT1X*, *GPX3*, *SPP1*, and *MXD1*). To know the prognostic value of these genes, we determined their correlation with prognosis by using the RNA-sequencing data and clinical information from 43 normal samples and 519 HNSCC samples released by TCGA (Figure 1a). First, we compared the expression level of common and valuable p-EMT-related genes in primary tumors and normal tissue. Higher expression of all common p-EMT-related genes were observed in primary tumors, compared to normal tissue (Figure 1b). Among 10 valuable p-EMT-related genes, *CXCL13*, *FN1*, *MMP3*, *MMP9*, *SPP1*, and *THBS2* were highly expressed in primary tumors (Figure 1b).

Next, we determined the correlation between the expression levels of p-EMT-related genes and prognosis in HNSCC cases by Log-rank test. We divided the cases into two groups, "low" and "high", based on the median expression of each p-EMT-related gene. Then, we compared the survival rate of "low" group with "high" group (Figure 2a and Supplementary Figure S1). Among these genes, *Serpin Family E Member 1* (*SERPINE1*), *transforming growth factor (TGF)-ß-induced* (*TGFBI*), *Integrin Subunit Alpha 5* (*ITGA5*), *cadherin 13* (*CDH13*), *Prolyl 4-Hydroxylase Subunit Alpha 2* (*P4HA2*), *Laminin Subunit Gamma 2* (*LAMC2*), and *Metallothionein 1X* (*MT1X*) were significantly correlated with poor prognosis (Figure 2a). These molecules besides *MT1X* were identified as a prognosis-related gene. As *MT1X* expression in HNSCC cases was lower than that in normal tissues (Figure 1b), *MT1X* was excluded. The expression patterns of these prognosis-related genes showed a positive correlation between 519 HNSCC samples obtained from TCGA (Figure 2b). In the gene expression cluster enriched p-EMT genes, *SNAI2*, which is involved in classical EMT program, was also included (Figure 2b).

To identify the HNSCC cell lines with p-EMT features, we comprehensively examined the expression of p-EMT-, EMT- and epithelial differentiation-related genes in HNSCC cell lines by qPCR. We used HSC2, HSC3, HSC4, HOC719PE, HOC719NE, HOC313, HOC621, and MSCC-inv1 cells in this analysis. As we expected, HOC719NE and HOC313 with EMT phenotype showed low expression level of

epithelial differentiation-related genes (Figure 3 and Supplementary Figure S2). These cells showed high expression level of p-EMT-related genes as well as EMT-related genes. HSC2, HSC4, and HOC621 cells showed high expression level of epithelial differentiation-related genes and low expression level of EMT- and p-EMT-related genes, suggesting that these cells showed epithelial phenotype. HOC719PE, HSC3, and MSCC-inv1 cells showed high expression level of EMT- and p-EMT-related genes, while they also showed high expression level of epithelial differentiation-related genes, suggesting that these cells showed p-EMT phenotype.

MSCC-inv1 cells were isolated from MSCC-1 cells by *in vitro* invasion assay (Figure 4a). We compared the expression of p-EMT-related genes, EMT-related genes and epithelial differentiation-related genes by using previous microarray data (Kudo et al, 2006). MSCC-inv1 showed higher expression level of p-EMT related genes, compared with parent MSCC-1 cells (Figure 4b).

4. DISCUSSION

In this study, we identified the prognosis-related genes, such as *SERPINE1*, *TGFBI*, *ITGA5*, *CDH13*, *P4HA2*, and *LAMC2* among p-EMT-related genes in HNSCC. GO enrichment analysis revealed that these prognosis-related genes are correlated with cell-substrate adhesion and angiogenesis (Supplemental Figure S3). Upregulation of *SERPINE1*, *TGFBI*, *ITGA5*, *CDH13*, *P4HA2*, and *LAMC2* may be involved in malignant behaviors via loss of cell adhesion and promoting angiogenesis. Indeed, there are several studies on the involvement of these molecules in cell adhesion and angiogenesis. PAI-1, encoded by *SERPINE1*, is an indicator of poor prognosis in different types of cancer by TCGA database analysis (Li et al, 2018). In HNSCC, high PAI-1 expression was associated with a higher rate of metastasis development and poor clinical outcome in HNSCC patient (Chin et al, 2005; Speleman et al, 2007; Pavón et al, 2015). Overexpression of PAI-1 accelerated HNSCC cell migration mediated by the activation of PI3K/AKT pathway (Pavón et al, 2015). Moreover, *PAI-1* is involved in angiogenesis in several types of cancer (Li et al, 2018). Interestingly, *PAI-1* is regarded as a mesenchymal marker and thoroughly confirmed to be a pivotal downstream effector on EMT induced by TGF-β in various cancer cell lines

(Omori et al, 2016; Chung et al, 2018; Wang et al, 2017). TGFBI is a 68 kDa matricellular protein and a member of the fasciclin domain containing protein family that also includes periostin (Skonier et al, 1992). Although many reports indicate that *TGFBI* functions as a tumor suppressor, there is also convincing data in the literature reporting a tumor-promoting role for *TGFBI*. In our previous study, periostin is identified as an invasion promoting factor in HNSCC (Kudo et al, 2006). Indeed, periostin expression was only observed in EMT-induced HNSCC cells. In previous study, *TGFBI* is considered as a p-EMT marker and *TGFBI* positive HNSCC cells is used as p-EMT^{high} population (Puram et al, 2017). Interestingly, p-EMT^{high} population showed higher invasiveness than p-EMT^{low} population. *ITGA5* and *P4HA2* are involved in HIF related hypoxia pathway, which causes tumor angiogenesis and hematogenous metastasis (Toss et al, 2018; Koike et al, 2004). *LAMC2* is involved in cellular migration and adhesion (Fukushima et al, 1998). *CDH13*, an atypical member of the cadherin family, has been associated with poorer prognosis in various carcinomas (Andreeva et al, 2010). *CDH13* is upregulated in blood vessels growing through tumors and promotes tumor neovascularization (Andreeva et al, 2010). Thus, previous reports revealed that p-EMT genes that we identified as a prognosis-related gene were involved in tumor progression (Figure 5). During p-EMT program, these molecules may cooperate each other for unfavorable behaviors in HNSCC. To clarify the detailed mechanism, further studies will be required.

In the previous studies, HNSCC cases are classified by gene expression profiles (Chung et al, 2004; Pavón et al, 2012; Walter et al, 2013; De Cecco et al, 2015). In all studies, basal group (epithelial phenotype) showed better prognosis than other groups. De Cecco et al. classified HNSCC cases as 6 groups (Cl1-Cl6); human papilloma virus (HPV)-like (Cl1), mesenchymal (Cl2), hypoxia associated (Cl3), inflammatory (Cl4), classical (Cl5), and immunoreactive (Cl6) by a meta-analysis approach (De Cecco et al, 2015). Among them, Cl2 and Cl3 show a more aggressive behavior. EMT-related genes were included in Cl2, common p-EMT-related genes were included in Cl3, Cl4, and Cl5. Chung et al. classified HNSCC cases as 4 groups (G1-G4). G3 and G4 showed poorer prognosis than G1 and G2.

Interestingly, p-EMT-related genes were included in G3 and G4. Epithelial differentiation-related genes were included in G1 and EMT-related genes were included in G2. Cumulative findings indicate that p-EMT genes are involved in poor prognosis.

EMT is not binary process. There are many cells in intermediate state of EMT, and the ability of proliferation, invasion and metastasis is different among each subpopulation (Pastushenko et al, 2018). Recently, it is recognized that the ability of cancer cells to undergo p-EMT, rather than complete EMT, poses a higher metastatic risk (Saitoh, 2018). qPCR analysis in HNSCC cell lines classified into three groups; epithelial phenotype, p-EMT phenotype, and EMT phenotype (Figure 3). This finding helps us to examine the nature of p-EMT and/or EMT in HNSCC. As the nature of p-EMT is still unknown, HOC719 PE, HSC3 and MSCC-inv1 cells can be used as a useful cell model for investigating a nature of p-EMT. Moreover, it is interesting to examine how the cells lose the expression of epithelial differentiation-related genes during EMT induction by using these cell lines. We think that this mechanism may be the key for understanding the involvement of p-EMT and/or EMT-related genes in malignant behaviors of HNSCC.

In conclusion, p-EMT program may be involved in poor prognosis of HNSCC. The p-EMT-related genes we identified in this study can be used for prognostic marker in HNSCC. However, it is still unknown how cancer cells require the phenotype of p-EMT during cancer progression. To understand the mechanism of p-EMT induction may lead to develop novel diagnosis and therapy for HNSCC.

CONFLICT OF INTEREST

The authors declare that they have no conflict of interests.

AUTHOR CONTRIBUTIONS

YK conceived and designed the study. SK, SW, NF, YM, TT, SJ and RA performed the research and statistical analysis. NI reviewed data and interpretation. SK and YK drafted the paper. RA and NI edited and revised the paper. All authors read and approved the final manuscript.

References

Aiello, N. M., Maddipati, R., Norgard, R. J., Balli, D., Li, J., Yuan, S., … Stanger, B. Z. (2018). EMT Subtype Influences Epithelial Plasticity and Mode of Cell Migration. *Developmental Cell,* 45(6), 681–695. doi: 10.1016/j.devcel.2018.05.027

Andreeva, A. V., & Kutuzov, M. A. Cadherin 13 in cancer. (2010). *Genes, Chromosomes and Cancer,* 49(9): 775-90. doi: 10.1002/gcc.20787

Bednarz-Knoll, N., Alix-Panabie` res, C., & Pantel, K. (2012). Plasticity of disseminating cancer cells in patients with epithelial malignancies. *Cancer and Metastasis Reviews,* 31(3-4), 673–687. doi: 10.1007/s10555-012-9370-z

Chin, D., Boyle, G. M., Williams, R. M., Ferguson, K., Pandeya, N., Pedley, J., … Coman, W. B. (2005). Novel markers for poor prognosis in head and neck cancer. *International Journal of Cancer,* 113(5), 789-797. doi: 10.1002/ijc.20608

Chung, C.H., Parker, J.S., Karaca, G., Wu, J., Funkhouser, W.K., Moore, D., … Perou, C.M. (2004). Molecular classification of head and neck squamous cell carcinomas using patterns of gene expression. *Cancer Cell*, 5(5), 489–500.

Chung, C. L., Wang, S. W., Martin, R., Knölker, H. J., Kao, Y. C., Lin, M. H., … Chen, C. L. (2018). Pentachloropseudilin inhibits transforming growth factor-β (TGF-β) activity by accelerating cell-surface type II TGF-β receptor turnover in target cells. *ChemBioChem,* 19(8), 851-864. doi: 10.1002/cbic.201700693

De Cecco, L., Nicolau, M., Giannoccaro, M., Daidone, M.G., Bossi, P., Locati, L., Licitra, L., Canevari, S. (2015). Head and neck cancer subtypes with biological and clinical relevance: Meta-analysis of gene-expression data. *Oncotarget*, 6(11), 9627-9642.

Fukushima, Y., Ohnishi, T., Arita, N., Hayakawa, T., & Sekiguchi, K. (1998). Integrin alpha3beta1-mediated interaction with laminin-5 stimulates adhesion, migration and invasion of malignant glioma cell. *International Journal of Cancer,* 76(1): 63-72. doi: 10.1002/(sici)1097-0215(19980330)76:1<63::aid-ijc11>3.0.co;2-h

Koike, T., Kimura, N., Miyazaki, K., Yabuta, T., Kumamoto, K., Takenoshita, S., … Kannagi, R. (2004). Hypoxia induces adhesion molecules on cancer cells: A missing link between Warburg effect and induction of selectin-ligand carbohydrates. *Proceedings of the National Academy of Sciences of the United*

States of America, 101(21), 8132-8137. doi: 10.1073/pnas.0402088101

Kudo, Y., Kitajima, S., Sato, S., Miyauchi, M., Ogawa, I., & Takata, T. Establishment of an oral squamous cell carcinoma cell line with high invasive and p27 degradation activities from a lymph node metastasis. Oral Oncol 2003; 39: 515-520.

Kudo, Y., Kitajima, S., Sato, S., Miyauchi, M., Ogawa, I., & Takata, T. (2004). Invasion and Metastasis of Oral Cancer Cells Require Methylation of E-Cadherin and/or Degradation of Membranous β-Catenin. *Clinical Cancer Research,* 10(16), 5455-5463. doi: 10.1158/1078-0432.CCR-04-0372

Kudo, Y., Ogawa, I., Kitajima, S., Kitagawa, M., Kawai, H., Gaffney, P. M., Miyauchi, M., & Takata, T. (2006). Periostin promotes invasion and anchorage-independent growth in the metastatic process of head and neck cancer. *Cancer Research,* 66(14), 6928-6935. doi: 10.1158/0008-5472.CAN-05-4540

Lee, B., Villarreal-Ponce, A., Fallahi, M., Ovadia, J., Sun, P., Yu, Q. C., … Dai, X. (2014). Transcriptional mechanisms link epithelial plasticity to adhesion and differentiation of epidermal progenitor cells. *Developmental Cell,* 29(1), 47–58. doi: 10.1016/j.devcel.2014.03.005

Li, S., Wei, X., He, J., Tian, X., Yuan, S., & Sun, L. (2018). Plasminogen activator inhibitor-1 in cancer research. *Biomedine & Pharmacotherapy,* 105, 83-94. doi: 10.1016/j.biopha.2018.05.119

Nguyen, P. T., Tsunematsu, T., Yanagisawa, S., Kudo, Y., Miyauchi, M., Kamata, N., & Takata, T. (2013). The FGFR1 inhibitor PD173074 induces mesenchymal-epithelial transition through the transcription factor AP-1. *British Journal of Cancer,* 109(8), 2248-2258. doi: 10.1038/bjc.2013.550

Nieto, M. A., Huang, R. Y., Jackson, R. A., & Thiery, J. P. (2016) EMT: 2016. *Cell,* 166(1), 21-45. doi: 10.1016/j.cell.2016.06.028.

Ocaña, O. H., Córcoles, R., Fabra, A., Moreno-Bueno, G., Acloque, H., Vega, S., … Nieto, M. A. (2012). Metastatic colonization requires the repression of the epithelial-mesenchymal transition inducer Prrx1. *Cancer Cell,* 22(6), 709–724. doi: 10.1016/j.ccr.2012.10.012.

Omori, K., Hattori, N., Senoo, T., Takayama, Y., Masuda, T., Nakashima, T., … Kohno, N. (2016). Inhibition of plasminogen activator inhibitor-1 attenuates

transforming growth factor-β-dependent epithelial mesenchymal transition and differentiation of fibroblasts to myofibroblasts. *PLoS One*, 11(2), e0148969. doi: 10.1371/journal.pone.0148969.

Pastushenko, I., Brisebarre, A., Sifrim, A., Fioramonti, M., Revenco, T., Boumahdi, S., … Blanpain, C. (2018). Identification of the tumour transition states occurring during EMT. *Nature,* 556(7702), 463-468. doi: 10.1038/s41586-018-0040-3.

Pavón, M.A., Parreño, M., Téllez-Gabriel, M., Sancho, F.J., López, M., Céspedes, M.V., … Mangues, R. (2012). Gene expression signatures and molecular markers associated with clinical outcome in locally advanced head and neck carcinoma. *Carcinogenesis*, 33(9), 1707-1716. doi: 10.1093/carcin/bgs207.

Pavón, M. A., Arroyosolera, I., Téllezgabriel, M., León, X., Virós, D., López, M., … Lópezpousa, A. (2015). Enhanced cell migration and apoptosis resistance may underlie the association between high SERPINE1 expression and poor outcome in head and neck carcinoma patients. *Oncotarget*, 6(30), 29016-29033. doi: 10.18632/oncotarget.5032.

Puram, S. V., Tirosh, I., Parikh, A. S., Patel, A. P., Yizhak, K., Gillespie, S., … Bernstein, B. E. (2017). Single-Cell Transcriptomic Analysis of Primary and Metastatic Tumor Ecosystems in Head and Neck Cancer. *Cell,* 171(7), 1611-1624. doi: 10.1016/j.cell.2017.10.044.

Saitoh, M. (2018). Involvement of partial EMT in cancer progression. *The Journal of Biochemistry*, 2018; 164(4), 257-264. doi: 10.1093/jb/mvy047.

Skonier, J., Neubauer, M., Madisen, L., Bennett, K., Plowman, G. D., & Purchio, A. F. (1992). cDNA cloning and sequence analysis of beta ig-h3, a novel gene induced in a human adenocarcinoma cell line after treatment with transforming growth factor-beta. *DNA and Cell Biology*, 11(7), 511–522.

Speleman, L., Kerrebijn, J. D., Look, M. P., Meeuwis, C. A., Foekens, J. A., & Berns, E. M. (2007). Prognostic value of plasminogen activator inhibitor-1 in head and neck squamous cell carcinoma. *Head & Neck*, 29(4), 341-350.

Tam, W. L., & Weinberg, R. A. (2013). The epigenetics of epithelial-mesenchymal plasticity in cancer. *Nature Medicine*, 19 (11), 1438–1449. doi: 10.1038/nm.3336.

Toss, M. S., Miligy, I. M., Gorringe, K. L., AlKawaz, A., Khout, H., Ellis, I. O., … Rakha, E.

A. (2018). Prolyl-4-hydroxylase A subunit 2 (P4HA2) expression is a predictor of poor outcome in breast ductal carcinoma in situ (DCIS). *British Journal of Cancer,* 119(12), 1518-1526. doi: 10.1038/s41416-018-0337-x.

Tsai, J. H., Donaher, J. L., Murphy, D. A., Chau, S., & Yang, J. (2012). Spatiotemporal regulation of epithelial-mesenchymal transition is essential for squamous cell carcinoma metastasis. *Cancer Cell,* 22(6), 725–736. doi: 10.1016/j.ccr.2012.09.022.

Walter, V., Yin, X., Wilkerson, M.D., Cabanski, C.R., Zhao, N., Du, Y., … Hayes, D.N. (2013). Molecular subtypes in head and neck cancer exhibit distinct patterns of chromosomal gain and loss of canonical cancer genes. *PLoS One*, 8(2), e56823. doi: 10.1371/journal.pone.0056823

Wang, X., Liu, C., Wang, J., Fan, Y., Wang, Z., & Wang, Y. (2017). Oxymatrine inhibits the migration of human colorectal carcinoma RKO cells via inhibition of PAI-1 and the TGF-β1/Smad signaling pathway. *Oncology Reports,* 37(2), 747-753. doi: 10.3892/or.2016.5292.

Yokoyama, K., Kamata, N., Hayashi, E., Hoteiya, T., Ueda, N., Fujimoto, R., & Nagayama, M. (2001). Reverse correlation of E-cadherin and snail expression in oral squamous cell carcinoma cells in vitro. *Oral Oncology,* 37(1), 65-71.

Figure legends

Figure 1. Study overview. (a) From TCGA-HNSCC RNA-seq data, we picked up the gene expression of p-EMT-related genes. To confirm the higher expression of p-EMT-related genes in HNSCCs, we compared the expression levels of HNSCC tissues with normal tissues. We also examined the correlation between p-EMT-related genes and prognosis of HNSCC patients. Then, we identified the p-EMT-related genes that are involved in poor survival; (b) p-EMT-related genes including common p-EMT and variable p-EMT genes are highly expressed in TCGA HNSCC tumors (n=519) compared to normal epithelium (n=43).

Figure 2. Correlation of the expression of p-EMT-related genes and poor prognosis. (a) Patients with high expression of p-EMT-related genes show poor disease-free survival. Red-high and blue-low expression of p-EMT-related genes as assessed from the TCGA data; (b) Heatmap shows the pairwise correlation in gene expression between 519 HNSCC samples obtained from TCGA. Correlation heatmap visualized three gene expression clusters. Each cluster enriched EMT-, p-EMT-, and epithelial differentiation-related genes were highlighted in red, yellow and blue, respectively. Purple circles highlighted the prognosis-related genes.

Figure 3. Expression of epithelial differentiation-, p-EMT-, and EMT-related genes in HNSCC cell lines. Heatmap shows epithelial differentiation-, p-EMT-, and EMT-related genes (rows) that are differentially expressed across HNSCC cell lines (columns). Purple circles highlighted the prognosis-related genes.

Figure 4. Expression of epithelial differentiation-, p-EMT-, and EMT-related genes in MSCC-1 and MSCC-inv1 cells. (a) Highly invasive clone, MSCC-inv1 cells were previously isolated from parent MSCC-1 cells by using *in vitro* invasion assay; (b) Graph shows the fold change of the expression of epithelial differentiation-, p-EMT-, and EMT-related genes in MSCC-inv1 cells. Among these genes, *PDPN, TNC, MMP2, VIM, CXCL13, FN1, MMP3, RAB25, MT1X, GPX3, EPCAM, KRT15, KRT6B, KRT6C, KRT17, KRT75, S100A7, S100A9, ZEB1, ZEB2, SNAI2,* and *SNAI1* were not expressed in both

MSCC-1 and MSCC-inv1 cells. Purple circles highlighted the prognosis-related genes.

Figure 5. Schematic model for involvement of prognosis-related p-EMT gene in unfavorable behaviors of HNSCC. *TGFBI*, *SERPINE1*, and *CDH1* inhibit cell-cell adhesion. *SERPINE1*, *CDH13*, *ITGA5*, *LAMC2* promote migration. *SERPINE1* promotes angiogenesis via HIF-1-VEGF axis. *CDH13* promotes angiogenesis. *ITGA5* promotes angiogenesis via PI3K-AKT signaling pathway. *SERPINE1*, *ITGA5*, *LAMC2*, and *P4HA2* are involved in extracellular matrix organization.

Figure 1

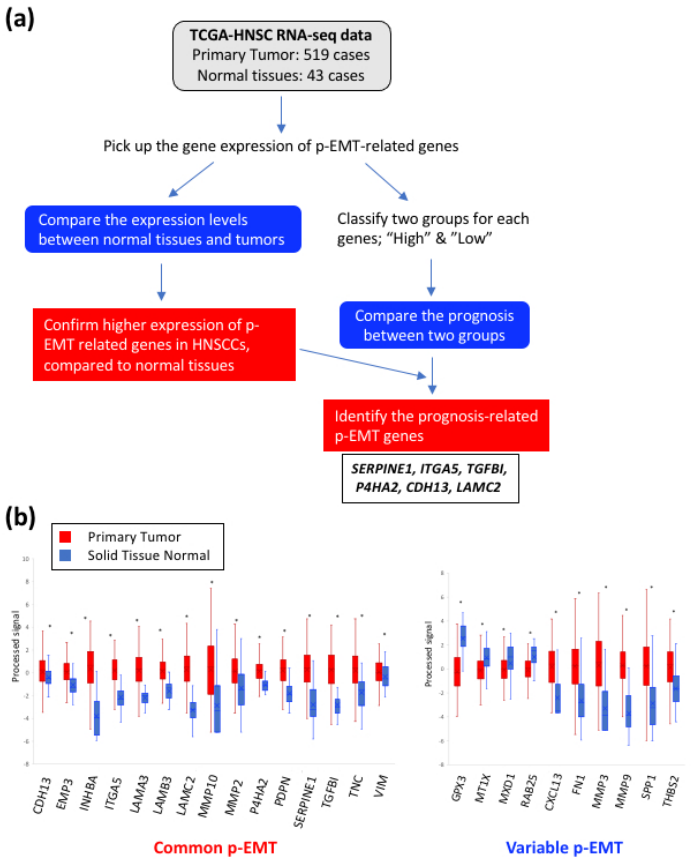


Figure 1

190x275mm (96 x 96 DPI)

Figure 2

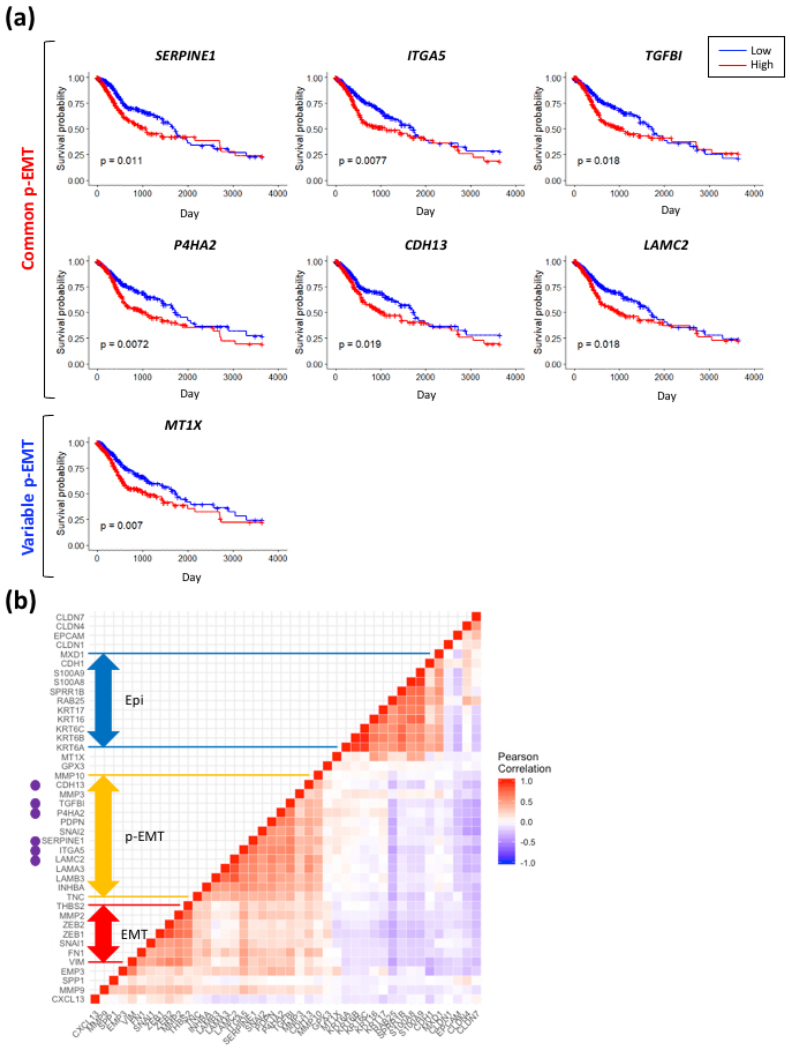


Figure 2

190x275mm (96 x 96 DPI)

Figure 3

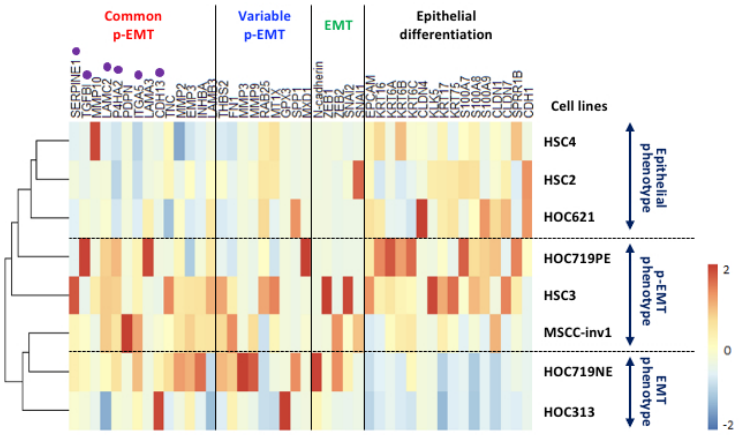


Figure 3

190x275mm (96 x 96 DPI)

Figure 4

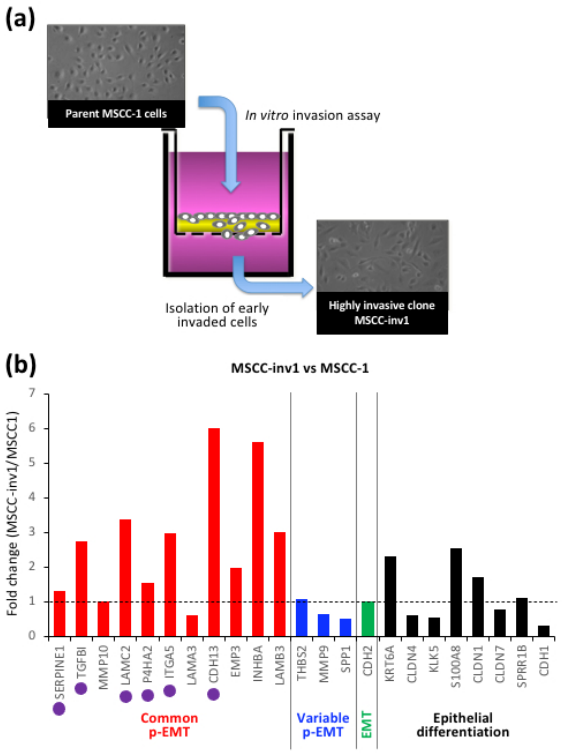


Figure 4

190x275mm (96 x 96 DPI)

Figure 5

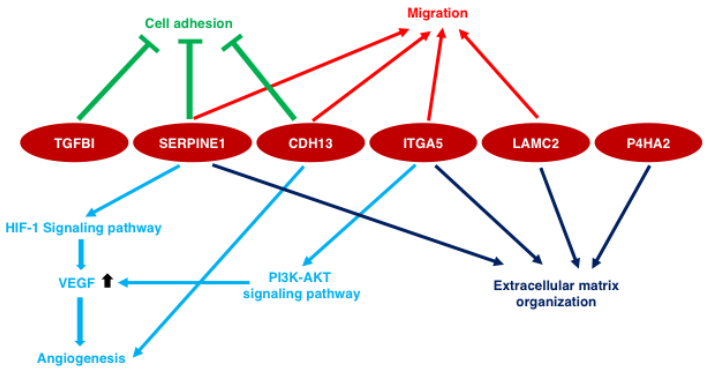


Figure 5

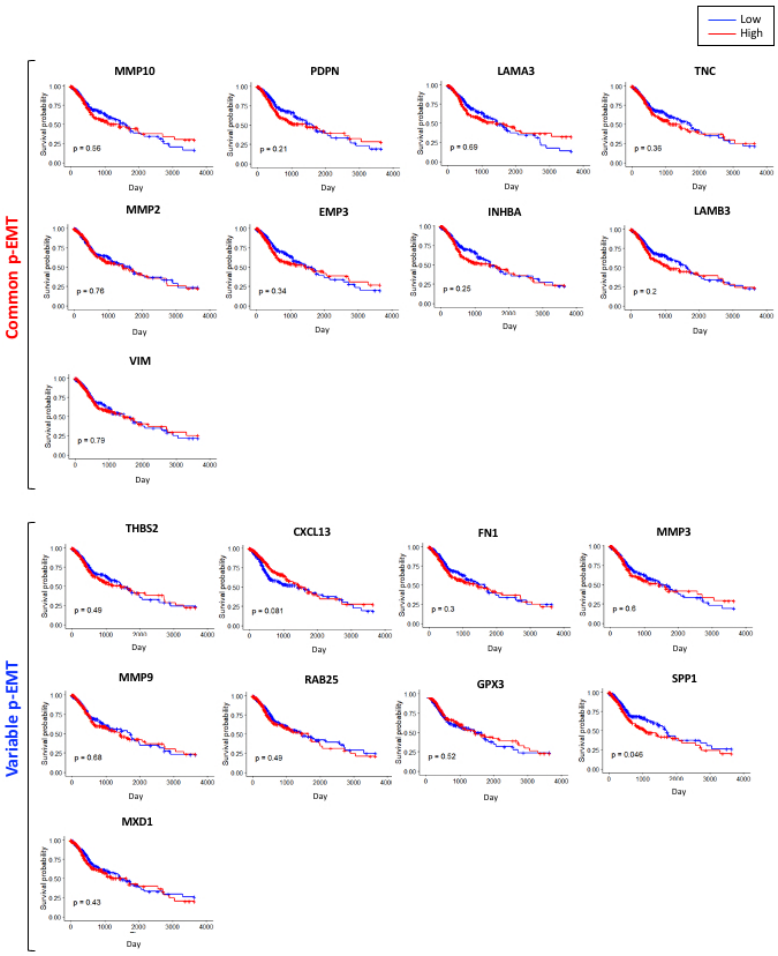
190x275mm (96 x 96 DPI)

Supplementary Table S1. List of primers used in this study

	Gene		genbank	Forward	Reverse
Common p-EMT genes	SERPINE1	serpin family E member 1	NM_000602.4	cgctgggaaagcattaagag	cacgcccagctaatttttgt
	TGFBI	transforming growth factor beta induced	NM_000358.3	gtgtgtgctgtgcagaaggt	ttgagagtggtagggctgct
	MMP10	matrix metallopeptidase 10	NM_002425.3	ggctcttcactcagccaac	tcccgaaggaacagatttttg
	LAMC2	laminin subunit gamma 2	NM_005562.2	gtcactggagaacgctgtga	agacccatttcgttgggacag
	P4HA2	prolyl 4-hydroxylase subunit alpha 2	NM_001017973.1	tgtcaaactgacacccctga	ggactcctgtcttgggatca
	PDPN	podoplanin	NM_001006624.1	catcgaggatctgccaactt	acgatgattgcaccaatgaa
	ITGA5	integrin subunit alpha 5	NM_002205.4	ctacaatgatgtggccatcg	ggatatccattgccatccag
	LAMA3	laminin subunit alpha 3	NM_000227.4	agatgaggcacatggagacc	ttctttttgcgcttttgtgttg
	CDH13	cadherin 13	NM_001220488.1	tgatgatgccaaaaacctca	atggtcagtttgtagttttgc
	TNC	tenascin C	NM_002160.3	ggtacagtgggacagcaggt	gttaacgccctgactgtggt
	MMP2	matrix metallopeptidase 2	NM_001127891.2	atgacagctgcaccactgag	atttgttgcccaggaaagtg
	EMP3	epithelial membrane protein 3	NM_001313905.1	gtggtctcagccctttcacat	atgagggagagcacatgtag
	INHBA	inhibin subunit beta A	NM_002192.3	cctcggagatcatcacgttt	ccctttaagcccactttctc
	LAMB3	laminin subunit beta 3	NM_000228.2	gggggagatcacaaactttga	gtgctggcagacacagacat
	VIM	vimentin	NM_003380.5	attgcaggaggagatgcttca	gtggagtttcttcaaaaaggca
Variable p-EMT genes	THBS2	thrombospondin 2	NM_003247.3	tcctgaaaacaatgccatca	gtccacacgaccaaactcgt
	CXCL13	C-X-C motif chemokine ligand 13	NM_006419.2	ctctgcttctcatgctgctg	tgagggtccacacacacaat
	FN1	fibronectin 1	NM_001306129.1	cagtgggagacctcgagaag	gtccctcggaacatcagaaa
	MMP3	matrix metallopeptidase 3	NM_002422.5	gcagttgctctcagcctatcc	gagtgtcggagtccagcttc
	MMP9	matrix metallopeptidase 9	NM_004994.2	ttgacagcgacaagaagtgg	gccattcacgtcgtccttat
	RAB25	RAB25, member RAS oncogene family	NM_020387.3	ccatcacctcggcgtactat	ttgttacccacgagcatga
	MT1X	metallothionein 1X	NM_005952.3	acccccaactgctcctgct	tctgacgtcccttgcagat
	GPX3	glutathione peroxidase 3	NM_001329790.1	tgcaaccaatttggaaaaca	ttcatgggttcccagaagag
	SPP1	secreted phosphoprotein 1	NM_000582.2	cccacagaccctttcaagta	ggggacaactggagtgaaaa
	MXD1	MAX dimerization protein 1 (MXD1)	NM_001202513.1	tgaacatggttatgcctcca	acttgattcgggtccaagtg
Epithelial differentiation	EPCAM	epithelial cell adhesion molecule	NM_002354.2	gctgtgtgtgtgaaacatgct	acgccttgtgatctccttct
	KRT16	keratin 16	NM_005557.3	tcttgctggatgtgaagacg	gaagctggatgagctctgct
	KRT6A	keratin 6A	NM_005554.3	ccaaggcagacactctcaca	gcagctctctgtactttgttc
	KRT6B	keratin 6B	NM_005555.3	caaggcccaatatgaggaga	ttggcacactgcttcttgac
	KRT6C	keratin 6C	NM_173086.4	ccaaggcagacactctcaca	gcagctctctgtactttgttc
	CLDN4	claudin 4	NM_001305.4	ctccatgggctacaggtaa	agcagcgagtcgtacacctt
	KLK5	kallikrein related peptidase 5	NM_001077491.1	agtcagaaaaggtgcgagga	tgaacttgcagaggttcgtg
	KRT17	keratin 17	NM_000422.2	gctgctacagcttttggctct	tcacctccagctcagtgttg
	KRT75	keratin 75	NM_004693.2	ggcatccaagaggtcactgt	agaggggctctaggttctgc
	S100A7	S100 calcium binding protein A7	NM_002963.4	tgctgacgatgatgaaggag	atgtctcccagcaaggacag
	S100A8	S100 calcium binding protein A8	NM_001319196.1	atgccgtctacagggatgac	acgcccatctttatcaccag
	S100A9	S100 calcium binding protein A9	NM_002965.4	cagctggaacgcaacatagg	tcagctgcttgtctgcattt
	CLDN1	claudin 1	NM_021101.4	ccgttggcatgaagtgtatg	ccagtgaagagagcctgacc
	CLDN7	claudin 7	NM_001185022.1	aattttcatcgtggcaggtc	aggacaggaacaggagagca
	SPRR1B	small proline rich protein 1B	NM_003125.2	cctccacctcaggaaccat	ttctgcttggtcttctgctg
	CDH1	cadherin 1	NM_001317184.1	tgcccagaaaatgaaaaagg	ggatgacacagcgctgtag
EMT genes	CDH2	cadherin 2	NM_007664.5	gacaatgccccctcaagttt	ccatttaagccgagtgatggt
	ZEB1	zinc finger E-box binding homeobox 1	NM_001128128.2	tgcactgagtgtggaaaagc	tggtgatgctgaaagagacg
	ZEB2	zinc finger E-box binding homeobox 2	NM_001171653.1	tgccaagagaggaagaggaa	gtgtcactgcgctgaaggta
	SNAI2	snail family transcriptional repressor 2	NM_003068.5	cttttttcttgccctcactgc	gcttcggagtgaagaaatgc
	SNAI1	snail family transcriptional repressor 1	NM_005985.4	tttacctccagcagccccta	cagtggggacaggagaagg

190x275mm (96 x 96 DPI)

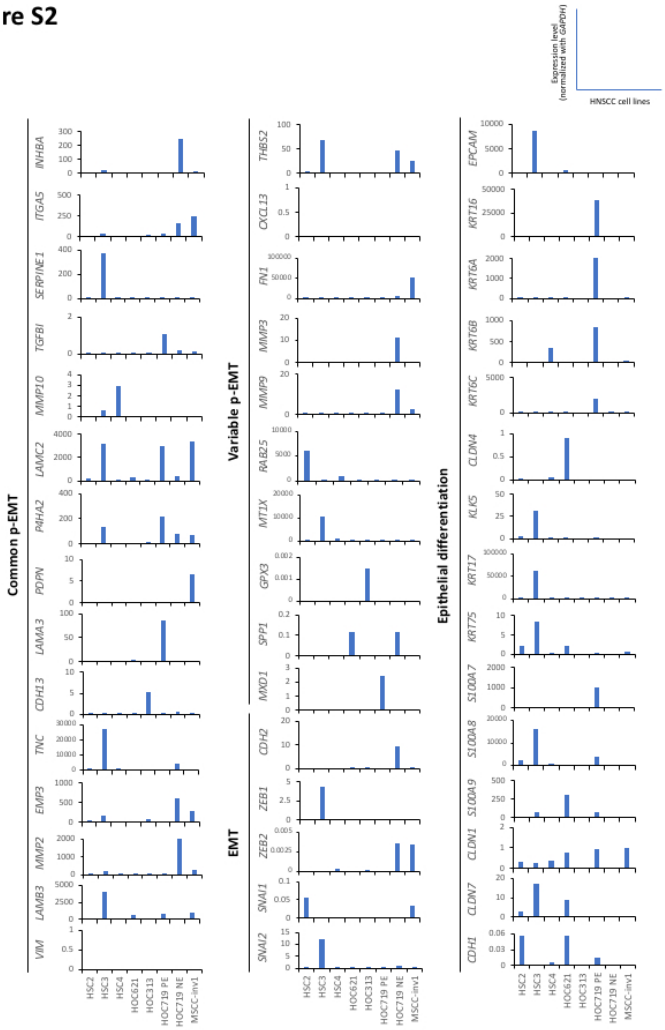
Figure S1



Supplemental Figure S1. Correlation between the expression of p-EMT-related genes and poor prognosis. Patients with high expression of p-EMT-related genes shown in these graphs does not significantly correlated with poor disease-free survival. Red-high and blue-low expression of p-EMT-related genes as assessed from the TCGA data.

190x275mm (96 x 96 DPI)

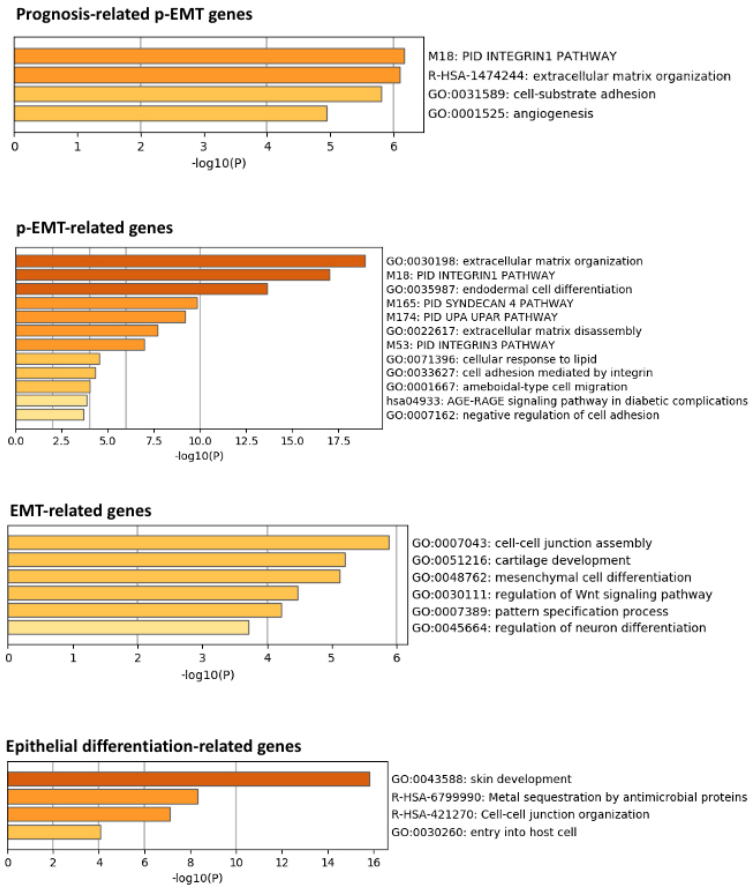
Figure S2



Supplemental Figure S2. Expression of epithelial differentiation-, p-EMT-, and EMT-related genes in HNSCC cell lines. We used HSC2, HSC3, HSC4, HOC621, HOC313, HOC719PE, HOC719NE, and MSCC-inv1. Graph shows the expression levels of each genes (normalized with GAPDH).

190x275mm (96 x 96 DPI)

Figure S3



Supplemental Figure S3. Bar graph colored by the statistical significance shows enriched gene ontology (GO) terms for prognosis related genes (p-EMT genes involved in poor prognosis), *SERPINE1*, *TGFBI*, *ITGA5*, *CDH13*, *P4HA2*, and *LAMC2*. Moreover, Enriched GO terms for p-EMT-, EMT-, and epithelial differentiation-related genes are also shown. GO analyses were performed in Metascape.

190x275mm (96 x 96 DPI)